The Impact of Real-Time Articulatory Information on Phonetic Transcription: Ultrasound-Aided Transcription in Cleft Lip and Palate Speech.

Joanne Cleland[1]*, Susan Lloyd[1&3], Linsay Campbell[2], Lisa Crampin[2], Juha-Pertti Palo[1,3], Eleanor Sugden[1], Alan Wrench[5], Natalia Zharkova[4]

[1]University of Strathclyde, Department of Speech and Language Therapy, Glasgow, UK
[2]NHS Greater Glasgow and Clyde, Dental Hospital and School, Glasgow, UK
[3]Queen Margaret University, Speech and Hearing Sciences, Edinburgh, UK.
[4]Newcastle University, Speech and Language Sciences, Newcastle, UK
[5] Articulate Instruments Ltd, Edinburgh, UK

Ultrasound-aided Transcription in Cleft Lip and Palate Speech

*Corresponding Author
Joanne Cleland
University of Strathclyde
School of Psychological Sciences and Health
Department of Speech and Language Therapy
Graham Hills Building
40 George Street
GLASGOW
G1 1QE

0141 548 3037
joanne.cleland@strath.ac.uk

Keywords: Cleft lip/palate; Speech; Speech and Language Therapy; Transcription; Ultrasound Tongue Imaging

Abstract

Objective: This study investigated whether adding an additional modality, namely ultrasound tongue imaging, to perception-based phonetic transcription impacted on the identification of compensatory articulations and on interrater reliability.

Patients and Methods: Thirty-nine English-speaking children aged 3 to 12 with cleft lip and palate (CLP) were recorded producing repetitions of /aCa/ for all places of articulation with simultaneous audio and probe-stabilised ultrasound. Three types of transcriptions were performed: 1. Descriptive observations from the live ultrasound by the clinician recording the data; 2. Ultrasound-aided transcription by two ultrasound-trained clinicians; and 3. Traditional phonetic transcription by two CLP specialists from audio recording. We compared the number of consonants identified as in error by each transcriber and then classified errors into eight different subcategories.

Results: Both the ultrasound-aided and traditional transcriptions yielded similar error-detection rates, however these were significantly higher than the observations recorded live in the clinic. Interrater reliability for the ultrasound transcribers was substantial (k=0.65), compared to moderate (k=0.47) for the traditional transcribers. Ultrasound-aided transcribers were more likely to identify covert errors such as double articulations and retroflexion than the audio-only transcribers.

Conclusion: Ultrasound-tongue imaging is a useful complement to traditional phonetic transcription for CLP speech.

Introduction

Transcribing speech is a key step in the decision-making process for children with speech sound disorders (SSD) [1]. A full and accurate phonetic transcription forms the foundation for differential diagnosis, treatment choices, measuring outcomes, and ultimately treatment success. However, phonetic transcription is known to suffer from issues with reliability, especially when the SSD is particularly severe and/or when the errors produced by the child involve phonetic distortions which result in speech sounds which are not expected with the phonological system of the target language [2]. For example, English-speaking transcribers find it particularly difficult to identify pharyngeal articulations [3]. Both severe SSD and these types of non-native productions can occur due to cleft lip and palate (CLP), which can make the speech of this group of children particularly challenging to transcribe [4, 5].

Cleft lip and/or palate (CLP) are the most common congenital craniofacial abnormalities, occurring in one in every 700 births [6] and problems with producing clear, intelligible speech can occur in CLP, even after successful surgery. Speakers with CLP often have difficulty achieving adequate velopharyngeal closure leading to hypernasality and nasal air emission. This in turn leads to a difficulty achieving adequate oral air pressure for high pressure consonants, leading to active compensatory articulations [7]. These compensatory articulations are characterized by retraction of anterior articulations often to sounds not occurring in the target language (at least in English, [8]). Moreover, a tendency for overuse of the tongue dorsum as a strategy to improve velopharyngeal closure [8] may lead to subtle phonetic distortions which might be difficult to transcribe.

In addition to speaker-related factors, such as the anatomical differences caused by CLP, transcription is also influenced by listener-related factors. Lack of familiarity with the target-language, or with the specific subtype of SSD might lead to less accurate transcriptions [9]. Likewise, categorical perception, which all listeners are subject to, can lead to transcription of phonemic

category collapses which might not truly represent the articulatory reality. Gibbon and Crampin [10] describe a case of an adult with CLP who produced both velar and alveolar targets as [c], however, instrumental analysis in fact revealed a subtle difference between /t/ and /k/ targets. Covert contrasts such as this are known to occur in the speech of both young typically developing children and older speakers with disordered speech [11] and are not readily identified with phonetic transcription.

Despite these difficulties, phonetic transcription by specialist listeners is still the "gold-standard" approach in CLP [12] and the approach used widely by speech and language therapists (SLTs) across the world. In part this is because instrumental techniques have, in the past, been impractical for use with young children [13] but also because perceptual analysis is important in its own right. Howard and Heslewood [14] argue that perceptual and instrumental analysis are "are two qualitatively different sides of the same phonetic coin" (p941). That is to say, that although instrumental techniques provide objective information on the movement of the articulators, only perceptual techniques such as phonetic transcription can truly represent the mental processes in the mind of the listener. This is important, because the ultimate goal of communication, and of course of speech therapy, is to be understood by the listener. Phonetic transcription also has the advantage that it is cheap, relatively quick, and requires only the skills of the SLT. Accuracy of these transcriptions can be improved in several ways: audio and/or video recording the speech samples for careful and repeated listening; use of listeners familiar with the client group; use of multiple transcribers [15]; acoustic analysis; or instrumental articulatory techniques. A recent study by Klintö and Lohmander [16] showed that using of video-recordings of the face, rather than audio-recording only, improved intertranscriber reliability and resulted in identification of more errors in three-year-olds with CLP. However, when age-appropriate phonological errors were removed from the analysis, the number of errors identified by audio-only versus audio plus video were not significantly different, suggesting that using video improves transcription only marginally. Acoustic analysis can also be used to

supplement phonetic transcription, and several studies have used various measures to identify covert contrasts (see for example [17] and [18]). A large body of literature, particularly in speakers with CLP, shows that instrumental articulatory techniques can be used to identify and quantify subtle phonetic errors in children's speech which might not be identified using transcription alone (see [19] for a list of electropalatography papers). Although instrumental techniques do not have the advantages of audio-only phonetic transcription in terms of being easy to use and inexpensive, often the instrumental technique in question is an obvious choice for remediating the child's SSD. For example, Cleland and colleagues [20] report an interesting case of a nine year old child "Rachel" who presented with a particularly persistent case of velar- fronting. Analysis with ultrasound tongue-imaging showed that Rachel presented with undifferentiated lingual gestures [21] and retroflexed productions of most stops. A follow-up paper [22] showed that Rachel was able to use ultrasound real-time visual biofeedback in therapy to achieve correct productions of velars. Thus, application of instrumental techniques to both assessment and intervention offers a dual benefit that transcription alone does not.

Instrumentation and Covert Errors in Cleft Lip and Palate

A number of speech errors which defy broad phonetic transcription have been reported in the literature, mainly in electropalatography (EPG) studies. EPG uses a custom-made pseudo-palate embedded with sensors (normally 62) to measure the timing and location of tongue-to-hard palate contact. A number of errors revealed by EPG are reviewed by Hardcastle and Gibbon [23], including misdirected articulatory gestures and double articulations where they ought not to exist. Despite potentially being missed by traditional phonetic transcription, these errors are important diagnostically because they provide evidence for articulatory, rather than phonological, difficulties, perhaps suggesting that different therapy approaches might be required. Because of this, and because of the potential for EPG to also be used as a biofeedback approach, the United Kingdom

Royal College of Speech and Language Therapists recommends EPG as an objective assessment and therapy approach for CLP.

Gibbon [24] summarises the EPG literature (23 papers over 20 years) on abnormal tongue-palate contact patterns in speakers with CLP. She suggests categorising errors into eight abnormal patterns: 1. increased contact; 2. retraction to palatal or velar articulation; 3. fronted placement; 4. complete closure (loss of grooving in sibilant productions); 5. open pattern (no tongue-to hard palate contact); 6. double articulations; 7. increased (phonetic-level) variability; and 8. Abnormal timing (e.g., articulatory groping). There is no discussion in Gibbon's paper [24] as to which of these errors might be vulnerable to being misidentified though audio-only transcription. Presumably errors such as retraction (i.e., classic "backing" in CLP speech where alveolars are produced at the velar place of articulation) are easy to transcribe when they result in native-speech sounds and category collapses, whereas errors such as double articulations and increased sub-phonemic variability will be harder to identify using transcription alone although this has not been empirically tested to date.

It is clear that using instrumentation such as EPG might add value to the transcription of disordered speech, although no previous studies directly compare audio-based transcription with articulatory-based transcription in a large number of children. In part this is because EPG is logistically difficult and expensive since each child requires a custom-made artificial palate and a period of stable dentition. To our knowledge EPG is never used for routine assessment in CLP due to costs and therefore studies of large numbers of children are lacking. In contrast, ultrasound tongue imaging (UTI), is becoming increasingly popular in the phonetics laboratory, but is relatively new to clinical phonetics.

Ultrasound tongue imaging.

Ultrasound Tongue Imaging (UTI) uses standard medical ultrasound to image the tongue in real-time, making it also suitable for visual biofeedback therapy. Over 30 small studies show it to be effective

for treating persistent SSDs (see for example [22] and [25]) and other studies use it for fine articulatory analysis of lingual movements when synchronised to the acoustic signal (for example, [20] or [26]). The ultrasound probe is placed under the chin, capturing most of the surface of the tongue in either the mid-sagittal or coronal plane. In both views, the imageable area is constrained by shadows from bone, with the tongue tip in particular being susceptible to a shadow from the mandible. Ultrasound has been used in a small number of studies to describe disordered speech. Both McAllister-Byun, Buchwald and Mizoguchi, [27] and Cleland and colleagues [20] used it to measure covert contrast in children with velar fronting. Cleland and colleagues [20] additionally used it to describe undifferentiated lingual gestures and retroflex articulations (see above). Cleland and colleagues compare phonetic transcription of velar stops in children with idiopathic SSD with ultrasound analysis and show that for the majority of children ultrasound analysis added very little information to the audio-transcription. The exception to this was "Rachel", where only careful ultrasound analysis revealed undifferentiated lingual gestures and retroflexion. Rachel was one of the children in this study with the most severe and complex SSD, suggesting that instrumental techniques may be most beneficial for this subgroup of children. However, ultrasound analysis has traditionally been a lab-based process requiring specialist software and time from a specialist speech scientist. An aim of the current study was therefore to determine whether observations from UTI could be realistically incorporated into the clinical environment.

In terms of using ultrasound to supplement phonetic transcription in CLP, Bressmann and colleagues [28] show covert articulatory movements during repetitions of /VkV/ in speakers with cleft palate, but no comparison with traditional transcription is given. Unlike Gibbon [24], Bressmann and colleagues make no attempt to classify the errors observed in ultrasound though they do note pharyngeals (which would be described as "open pattern" by Gibbon); fronted placement; and double articulations. In this study the observations from UTI are descriptive in nature, performed off-line by specialist researchers. While this is more time consuming that live phonetic transcription,

it may be analogous to the blinded specialist listener paradigm suggested as the gold standard by Sell [12].

1.3 Purpose and hypotheses

The purpose of the current study was to compare audio-only transcription of disordered speech to transcription accompanied by ultrasound tongue imaging. As discussed above, cleft-palate speech makes an excellent test case for this experiment as it is known to be vulnerable to subtle phonetic errors and non-native sounding productions; we therefore predict that ultrasound-based transcription will have an advantage in this client group as it allows visualisation of non-native speech sounds such as pharyngeals (in English-speaking children) and can reveal covert articulations. We hypothesise that an advantage might be demonstrated by: 1. An increase in the number of active compensatory errors identified when using ultrasound-aided transcription (UA) compared to audio-only transcription (AO) and 2. Increased interrater reliability when using UA compared to AO. A secondary aim of this study was to develop a clinician-friendly UA recording format for describing speech errors using ultrasound. We aimed to determine whether it was possible to identify ultrasound-based errors in real-time, live in the clinic (to emulate an SLT doing a live transcription) or whether off-line careful viewing of the ultrasound was necessary (similar to off-line analysis of video/audio recordings as is standard for CLP speech). We predicted that off-line careful viewing of ultrasound (UA) would have an advantage over live in-clinic transcription (CT) and that an advantage might be demonstrated by: 1. An increase in the number of compensatory errors identified when using UA compared to CT, and 2. Increased interrater reliability when using UA compared with CT.

A further exploratory aim of this study (to be quantified using ultrasound indices elsewhere) was to classify the errors according to Gibbon [24] and Cleland's [20] error types and to explore whether different error-types were more prevalent in ultrasound-aided transcriptions. That is, whether covert errors such as double articulations would be noted with increased frequency compared to audio-only transcriptions.

Method

Speakers

Children attending routine appointments over a 12 month period at the West of Scotland Cleft Lip and Palate Service were invited to participate. Inclusion criteria were: syndromic or non-syndromic CLP; aged 3 to 15, and spoken English. Both children with and without overt SSDs were included. Children with cleft lip only; severe learning disability; or no speech were excluded.

Thirty-nine children consented to taking part in the project. Of this, data from 35 children (15 female, 20 male) aged 3;07-12;02 (mean age = 6;09) was included for transcription. Three datasets were unusable due to very poor quality ultrasound images; one dataset was collected after files had been submitted for transcription. See Table 1 for biographical and medical information of the speakers.

Table 1: Biographical, medical, and language information for speakers with CLP

Participant number	Sex	Age (Years; months)	Cleft type	Additional Medical Diagnoses	Language Spoken at Home
1	M	10;05	BCLP	None	English
3	F	5;01	UCLP	None	English
5	M	10;03	UCLP	None	English
6	M	9;08	UCLP	None	English
7	M	9;02	UCLP	None	English, Spanish
8	M	4;07	UCLP	None	English
9	M	9;08	UCLP	Cluttering/stuttering	English
10	M	7;07	UCLP	None	English
11	M	4;05	CP	None	English
12	F	5;01	CP	Sticklers Syndrome	English
13	F	5;02	CP	None	Gaelic, English
14	F	5;10	BCLP	None	English
15	F	4;09	UCLP	90% of tonsils and 90% of adenoids removed at age 2	English
16	M	9;07	UCLP	None	English
17	F	4;04	BCLP	None	English
18	F	5;11	UCLP	None	English
19	F	3;09	CP	Treacher Collins Syndrome	English
20	F	7;05	CP	None	English
21	M	9;11	CP	None	English
22	M	10;01	CP	None	English
23	F	12;02	BCLP	None	English
24	M	6;05	CP	Sticklers Syndrome, micrognathia	English
25	M	4;11	CP	None	English
26	M	4;04	BCLP	None	English, Turkish

Materials

Speech materials were adapted from the CLEFTNET protocol originally developed for EPG recordings [29]. This comprised: 1. counting from one to 10; 2. Ten repetitions of all voiceless (or voiced where necessary) obstruents and sonorants in /aCa/; 3. sentences from GOS.SP.ASS 98 [30]; and 4. Five minimal (pair) sets containing contrasting common substitutions for /s, ʃ, tʃ, t/ in a variety of vowel environments (for example, “sheet, seat, cheat team, keep”) . Speech materials were presented orthographically on a laptop screen and pre-recorded audio prompts were provided for imitation. Younger children sometimes also required support from the researcher to imitate the prompts, for example, a further live prompt. All materials were collected by a Speech and Language Therapist trained in ultrasound data acquisition (the second author). Only the repetition data of single consonants were included for analysis in this paper.

Ultrasound Recording Set-up

High-speed ultrasound data were acquired using a Micro machine controlled via a laptop running Articulate Assistant Advanced software™ version 2.17 [31] which recorded both audio and ultrasound. The echo return data were recorded at ~100fps over a field of view of either 144 or 162 degrees. The field of view was selected with the probe positioned to allow the greatest view of the tongue, including both the hyoid and mandible shadows. The microconvex ultrasound probe was stabilised with a custom-made lightweight flexible plastic head set. For six children the probe was held in place by the hand, either as the head set was too big, or as the children requested not to use the head set. This may have affected the quality of the images. Data were collected in a quiet room at the Glasgow Dental Hospital before or after a routine appointment.

Data were collected first in the mid-sagittal view for all materials. The probe was then turned 90 degrees and the materials containing sibilants were collected again. In this coronal view it is possible

to see lateral contact and/or bracing which is important for /s, ʃ, t͡ʃ/. These data were subsequently excluded from the analysis due to poor image quality in 11 of the children. From a practical perspective it was very difficult to determine which coronal slice of the tongue (i.e., more anterior or posterior) was imaged in each child. Participants were asked to swallow a sip of water at the beginning and end of recording in each position to allow a palate trace to be drawn for future quantitative analyses.

An Audio-Technica 3350 microphone was attached to the headset (or child's clothes where the head set was not used) to record audio information simultaneously with the ultrasound recording.

Transcriptions

Three types of transcriptions were performed:

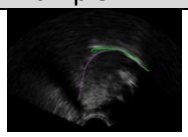
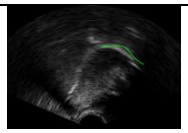
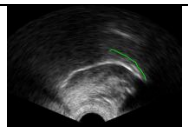
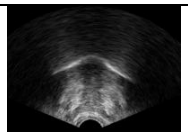
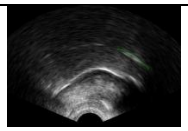
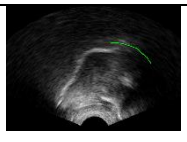
1. Descriptive observations live in the clinic by the SLT (the second author) recording the data, online during the recording session (CT)
2. Descriptive ultrasound-aided transcription by two ultrasound-trained SLTs (the first and sixth author hereafter UA1 and UA2) after the sessions (UA)
3. Traditional blinded phonetic transcription by two CLP specialist SLTs (AO1 and AO2) using the audio recording only (AO).

Live Transcriptions/Observations (CT)

An assessment form to record phonetic transcription and initial subjective impressions from the ultrasound was developed to record observations from the live ultrasound by the clinician recording the data (see [32]). Error-types were based on Gibbon [24] (note, we were unable to provide data on "complete closure" due to poor image quality in the coronal view, the description of the error type is included in table 2 for completeness) with the addition of retroflexion [20] which has not previously been identified using EPG but is highly salient on ultrasound. Since ultrasound shows tongue position

and shape, rather than tongue-palate contact as in EPG, we re-described the errors to better reflect what might be viewed using ultrasound. For example, “Open Pattern” occurs in EPG when there is no tongue-palate contact (a completely “white” EPG frame), this might be due to a post-velar production, e.g., a uvular production, or due to undershoot of an articulatory gesture. Using ultrasound it is possible to categorise both of these types of errors separately. The descriptions, along with example videos, are defined in an open-access manual [32] and summarised in table 2. The video examples were used to train the SLT collecting the data (CT) and later the off-line transcribers (UA1 and 2) in ultrasound error-types.

Table 2: EPG Error types, ultrasound equivalents, and expected (Ext)IPA transcriptions; green line denotes palate

Error Type	EPG Description*	Ultrasound Description	Ultrasound Example	Expected IPA Transcription
1. Increased Contact	Increased contact between the tongue and the hard palate	Raising of tongue body and tip/blade towards or in contact with the hard palate		Simultaneous alveolar + postalveolar + palatal
2. Retraction	Alveolar target retracted to velar or palatal	Alveolar target retracted to velar or palatal		Velar or palatal
3. Fronted	Posterior target is fronted to palatal, post-alveolar, or alveolar	Posterior target is fronted to palatal, post-alveolar, or alveolar		Alveolar, post-alveolar, or palatal
4. Complete Closure	Complete closure in the alveolar rows during sibilant production	No visible groove in the coronal view.		Any lateral sibilant
5. Open Pattern	No contact between tongue and hard palate	Uvular or pharyngeal articulation OR undershoot		Uvualar, pharyngal or "lowered" diacritic
6. Double Articulation	Simultaneous production of two consonants: normally alveolar-velar, but also lingual-labial	Simultaneous production of two consonants: normally alveolar-velar		Any double articulation e.g. [k͡t] or [p͡t]
7. Increased Variability	Different EPG patterns per repetition	Different tongue-shapes per repetition	*(dynamic analysis required)*	Different transcriptions across repetitions
8. Abnormal Timing	Mis-directed articulatory gestures or release of articulations with abnormal timing	Mis-directed articulatory gestures or release of articulations with abnormal timing	*(dynamic analysis required)*	Any diacritic denoting timing such as lengthening marks
9. Retroflexion	*NA*	Tongue tip retroflexion during any non-retroflex target		Any retroflex consonant

*EPG Descriptions are summarised from [23]

For the CT transcriptions, the second author recorded the data, marking observations from both the ultrasound and the live-productions of child's speech on the assessment form at the same time. This process emulated a traditional in-clinic live transcription which, as identified previously, is known to be problematic for CLP speech. CT transcribed errors using the International Phonetic Alphabet (IPA, [33] and the Extensions to the IPA (ExtIPA, [34]) and then classified the errors as: 1. Correct; 2. Classified into one or more of the above nine error types, or; 3. Classified as a non-imageable error, for example errors of voicing or passive cleft errors such as nasal escape. It was possible for an error to fall into multiple categories, for example alveolar targets might be retracted, variable, and with nasal escape.

Ultrasound-Aided Transcription (UA)

UA1 and UA2 transcribed the children's speech off-line. Prior to watching/listening to the simultaneous audio and ultrasound recordings, both transcribers watched the video exemplars of the error types [30]. Both transcribers were SLTs with a speciality in SSD (though not specifically cleft-palate speech), UA1 is an expert in ultrasound analysis of disordered speech and was responsible for training both UA2 and CT. Although it would have been desirable to train cleft-palate specialists to be proficient users of ultrasound it was not practical due to training-time constraints.

Both UA transcribers worked independently, but in the same room, to transcribe the children's speech. Multiple sessions of one to two hours each were required to transcribe all of the data. The transcribers watched (and listened to) each ultrasound recording only once, in real-time. The time to record observations after each viewing was not constrained. Errors were transcribed using IPA and ExtIPA symbols and then classified in the same way as the CT transcriptions, that is, as correct or into one or more of the 9 error types identified in Table 3, or as a non-imageable error.

Traditional Phonetic Transcription (AO)

Two specialist SLTs in CLP transcribed the data using symbols of the IPA and ExtIPA. Both clinicians had over ten years' experience in the transcription of cleft palate speech and were not associated with the research project. The audio signal was extracted from the ultrasound plus audio recordings of the children's speech and the AO transcribers were provided with .wav files only, it was therefore not possible for them to view any ultrasound.

A transcription guide was given to the AO transcribers detailing the contents of the recordings, and instructions to listen once to each file, then note the transcription, then move to the next recording. They were instructed to transcribe only the consonant for the VCV prompts, and were permitted to note only one transcription if they thought all 10 repetitions sounded the same. Following this, they were permitted to listen to each file again up to a maximum of three additional times and make any further transcriptions in an additional column. These were not included in the current study.

The AO transcriptions were then coded by the research team as either "correct" i.e. the transcription matched the target; classified into one of the above nine error types (see table 2 for expected phonetic transcriptions); or classified as a non-imageable error (i.e. a non-lingual error that cannot be viewed on ultrasound). This process was similar to a traditional phonetic/phonological pattern analysis.

Statistical Analyses

To determine whether use of ultrasound led to an increased error-identification rate we averaged the mean correct attempts within groups of transcribers (CT, UA, AO) and compared these using Friedman's Two-way analysis of variance by ranks.

To determine whether ultrasound-aided transcription led to improved interrater reliability we calculated Cohen's kappa with a 95% confidence interval between raters for three different subclasses of errors:

1. Correct verses incorrect productions only
2. Imageable (one of the nine ultrasound-visible error types) verses non-imageable errors (such as changes in manner of articulation or nasal resonance)
3. Interrater reliability across all nine error-types

For the purposes of interrater reliability only the first error identified (see table 2) was coded as Cohen's kappa does not allow multiple categorisation of a single error.

Descriptive Analysis of Error-Types

To determine whether the CT, UA, and AO transcribers were more likely to identify specific types of errors (for example whether errors which are traditionally thought of as covert were more likely to be noted in the ultrasound) visual inspection of the data was employed due to a small number of raters and speakers.

Results

Error Identification

There were a total of 770 items produced by the children and young people recorded. The CT transcriber annotated only 493 items whereas the UA and AO raters were able to transcribe between 629 and 712 items. The CT and UA transcribers were unable to detect presence or absence of complete closure (a "domed" rather than "butterfly-wing" shape in the coronal view) during sibilants due to poor image quality and these items were therefore excluded for all transcribers (see above). The CT transcriber identified 380/493 items as correct; UA transcribers identified on average

415/630 items as being correct and the AO transcribers identified an average of 432/707 items as correct.

A Friedman test identified no significant difference between all groups of transcribers in terms of the number of items identified as being "correct" $\chi^2(2)$ 2.217, p=.330. However, a Friedman test to compare the number of errors identified was significant, $\chi^2(2)$ 21.317, p<0.001. Dunn-Bonferroni post hoc tests showed that there was a significant difference between AO and CT (p<.0001) and also between UA and CT (p<.0001). In summary, our hypothesis that UA transcription would lead to identification of more active compensatory articulations than AO was not supported however there was evidence that UA and AO led to identification of more errors than live transcription (CT). This appears to be due to CT having a tendency to mark items as correct or leave them out, probably due to time pressure and ability to manage both the child's attention and the ultrasound recording. We would therefore not recommend that clinician's attempt live transcription/assessment using ultrasound.

Interrater Reliability

Table 3 summarises interrater agreement, calculated using Cohen's kappa with a 95% confidence interval. When classifying transcribed productions only as either correct or incorrect, agreement within transcription conditions was found to be "substantial" for UA transcription (κ = 0.65), and "moderate" (κ = 0.47) for AO transcribers. When comparing across transcription conditions, agreement was lower, either "fair" or "moderate" (κ ranging from 0.36-0.45).

Agreement for the CT transcriptions carried out at the time of recording was found to be "fair" (κ = 0.35 and 0.37) when compared with UA, and "moderate" when compared with AO (κ = 0.46 and 0.41).

Table 3. Interrater agreement calculated using Cohen's kappa with a 95% confidence interval

Pair compared	Correct/incorrect only		Imageable/Non-Imageable		All errors	
	K	Descriptor	*K*	Descriptor	*K*	Descriptor
Within Modality						
UA1/UA2	0.74	**substantial**	0.70	**substantial**	0.65	**substantial**
AO1/AO2	0.58	moderate	0.48	moderate	0.47	moderate
Across Transcriber Category						
UA1/AO1	0.40	fair	0.35	fair	0.31	fair
UA1/AO2	0.36	fair	0.29	fair	0.24	fair
UA2/AO1	0.45	moderate	0.40	fair	0.33	fair
UA2/AO2	0.36	fair	0.31	fair	0.24	fair
CT/UA1	0.35	fair	0.34	fair	0.30	fair
CT/UA2	0.37	fair	0.32	fair	0.25	fair
CT/AO1	0.46	moderate	0.36	fair	0.36	fair
CT/AO2	0.41	moderate	0.32	fair	0.30	fair

Table Note: UA ultrasound-aided condition; AO audio only condition; CT in-clinic transcription at time of data collection. **Substantial** agreement is in bold type.

The same relationship was found between the transcription conditions when errors were sub-classified into those that can (one ultrasound error-types) and cannot be imaged (a non-ultrasound error), and also when further sub-classified into the remaining eight error types, but less agreement overall was noted (see table 3). The interrater agreement for UA remained "substantial" (κ = 0.65) whether measuring incorrect/correct only, imageable/non-imageable errors, or classifying into the eight error types, whereas AO was moderate for all comparisons. In summary, UTI appears to lead to substantial interrater reliability for detecting and classifying lingual errors in children with CLP and in general has "fair" agreement with traditional audio-only transcriptions.

Descriptive Analysis of Error Types.

Figure 1 shows error-type classification by transcriber-group. Classifications from UA and AO were averaged across both transcribers in each pair as reliability was substantial or moderate. No

transcribers noted abnormal timing, this is probably due to difficulties noting this in real-time with further quantitative articulatory analysis required. The UA group noted substantially more instances of increased contact, double articulation, and retroflex productions, which were either not noted or, in the case of double articulation, noted in only one instance by the other transcribers. This may suggest a benefit of UTI in detecting these covert error types as predicted, however further work with more transcribers and more speakers is required. The AO group recorded higher numbers of retracted placement than both the UA and CT groups. Increased variability was noted by both UA and AO groups, but with higher rates in the AO transcribers. CT had a general pattern of small error-detection rates (see above), although the she did note non-imageable errors (such as nasalisation or voicing errors) frequently.

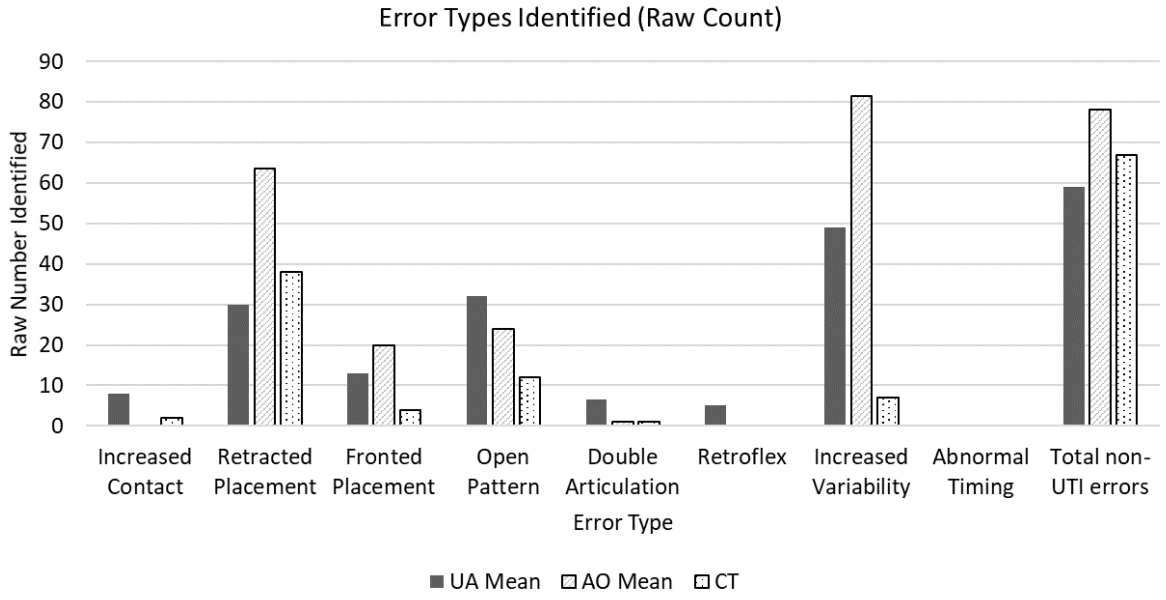


Fig. 1. Error types identified by transcriber group

Discussion

Through this study we sought to determine if adding real-time articulatory information in the form of ultrasound tongue-imaging has an impact on phonetic transcription. Shriberg and Lof [35] argue for the inclusion of instrumental approaches in analysing disordered speech and as a method for

circumventing some of the problems with phonetic-transcription. Here we take a different approach, using ultrasound as an additional modality *during* transcription, rather than presenting lab-based quantitative analysis in competition with perceptual approaches. We chose to investigate CLP speech as it is particularly vulnerable to problems with interrater reliability due to covert errors and non-native phonetic realisations, and because it has been previously studied with a similar instrumental technique, namely electropalatography. In line with Klintö and Lohmander [16] adding an extra modality, in our case ultrasound, had little impact on the percentage of consonants recorded as correct. Both the audio-only (AO) and ultrasound-aided (UA) transcribers noted similar rates of both correct and in-error productions suggesting that both methods are valid for making decisions about overall severity of SSD or potentially for measuring outcomes after intervention. While this was previously well-established for traditional transcription [13], to our knowledge this is the first group study to explicitly test the effect of adding an articulatory technique which provides information about internal articulations.

In contrast, an increased number of errors were identified when we compared live, in clinic, (CT) transcription with transcription performed offline, after the recording session. From this we conclude that using ultrasound to assess children "on-the-fly" may be unwise. Difficulties with noting errors may have been due to the transcriber's shifting attention between the child, the recording equipment, and the transcription, however further studies comparing different people transcribing live in the clinic are required.

In terms of interrater reliability, we found that within category of transcribers the UA pair had an advantage over the AO transcribers. Reliability was "moderate" for AO and "substantial" for UA. Thus, our hypothesis that using UA transcription may lead to increased interrater reliability compared to AO transcription was confirmed. While this effect remains to be tested across larger numbers of transcribers and larger numbers of speakers, it is promising preliminary evidence that instrumental analysis adds an objectivity to transcription [14] that can complement perceptual

transcription. Across-category, reliability between AO and UA was fair or moderate, even when looking at reliability across the different sub-categories of error. This reduction in reliability speaks to differences in the types of error identified by transcribers using ultrasound and transcribers using audio-only.

The UA transcribers classified errors into all of the types described by Gibbon [24] with the exception of complete closure and abnormal timing. Identifying complete closure was problematic due to difficulties with the coronal ultrasound view (though it should be noted that both transcribers commented that they could hear lateralisation at times). Further methodological work should seek to improve the positioning of the probe for coronal recordings. Although we did not identify abnormal timing it should be noted that EPG studies which have reported this phenomenon have relied on careful quantitative measurements of dynamic EPG patterns, rather than real-time viewings as performed here. Ongoing work seeks to use a range of quantitative ultrasound measures to determine whether abnormal timing can be identified in this group of children. Nevertheless, the UA transcribers did note increased rates of double articulations and "open-pattern" (undershoot or uvular/pharyngeal articulations) in the children's speech. Moreover, the transcribers noted a small number of instances of retroflex productions. This error has not previously been reported using instrumental analysis of CLP speech, although Cleland and colleagues, [20] do report retroflex articulations in a child with a severe idiopathic SSD. Thus, it appears that UA transcription can lead to the identification of unusual error-types, which may be more common in children with severe SSD or with structural differences such as CLP influencing their articulation.

Limitations

In line with Howard and Heselwood's [14] view that instrumental techniques should complement perceptual analysis, it is worth noting some of the limitations of the current study regarding both the study design and the use of ultrasound more generally. Firstly, our classification of errors was based

on Gibbon's classification system and was limited to those errors which involve tongue-shape; however, CLP speech is also highly vulnerable to difficulties with the velopharyngeal mechanism. Thus, these may not have been identified in this study.

We make no attempt here to determine whether the AO transcribers, who were experts in CLP, noted more instances of passive errors such as nasal escape although all groups of transcribers did identify these types of errors. In this sense ultrasound can only complement perceptual analysis, and not replace it. This remains the case even if quantitative ultrasound measures are used to provide a more objective error analysis than the one provided here. Nevertheless, ultrasound did enable identification of different error types which are potentially diagnostically important. Moreover, ultrasound, when used in the form of biofeedback, offers a potential approach for remediating these very errors [21].

The current study was constrained by the small number of transcribers. While it was necessary to have only one transcriber during the recording process (CT), further transcribers performing the UA and AO transcriptions would have been beneficial. Additionally, our UA and AO transcribers differed not only in the modality they were using, but also their own previous experience: both AO transcribers were specialists in CLP, whereas both UA transcribers were specialists in other types of SSD. A potential solution would be to repeat the experiment with CLP specialists trained in the use of ultrasound, time constraints prevented this in the current study. It would be equally useful to repeat the experiment with audio and ultrasound from speakers with different types of SSD.

Lastly, the study was constrained by the use of single viewings of real-time ultrasound only. While this gives the study potential ecological validity it is probable that performing careful articulatory analysis, including quantitative ultrasound measures, would provide different results, this work is currently ongoing. Moreover, the speech materials used here were deliberately constrained to multiple repetitions of single consonants in /aCa/ contexts. It is our view that making judgements

about tongue-shapes in real-time beyond single segment level is likely to be extremely difficult, although this remains to be empirically tested. We recommend play-back of more complex speech materials in slow motion or frame-by-frame, as well as quantitative analyses.

Conclusions

The addition of ultrasound tongue images to audio-perceptual information appears to improve the reliability of lingual error identification in cleft palate speech. Contrary to our hypothesis that the addition of ultrasound would lead to an increased number of errors being identified, we found no difference in the overall number of errors between audio-only and ultrasound-aided transcriptions. However, transcribers who were given additional ultrasound information were able to identify increased instances of double articulations, pharyngeal/uvular articulations and retroflexion within those productions identified as incorrect. However, it was not possible to detect these errors reliably in real-time in the clinic. Through this study we provide preliminary evidence that ultrasound may be a useful addition to the CLP assessment toolkit when images are recorded for later playback. This might be especially useful when ultrasound assessment is used as a precursor to ultrasound biofeedback therapy when the same technique can be used to remediate the very errors it helps identify.

Acknowledgements

We wish to thank all the children and their carers who gave up their valuable time to take part in this research project. Thank you to the Glasgow Dental Hospital and School for providing clinic space to make the recordings. We would also like to thank Stephanie van Eeden and Caroline Hattee for their transcription of the data, and David Young for his support with the statistical analysis.

Statement of Ethics

Participants and their parents/carers gave their written informed consent. The study protocol was approved by the National Health Service West of Scotland Research Ethics Committee and the University of Strathclyde Research Ethics Committee.

Disclosure Statement

The authors have no conflicts of interest to declare.

Funding Sources

This work was funded by a grant from Action Medical Research, GN2544.

Author Contributions

Cleland and Crampin designed this study and received the funding for it. Lloyd and Campbell collected the data. Lloyd, Cleland, and Sugden analysed the data. All authors were involved in early preparation if this work, Cleland and Lloyd wrote the final paper. All authors read and commented on a final draft of the paper.

References

1. Child Speech Disorder Research Network: Good Practice Guidelines for the Tanscription of Children's Speech Samples in Clinical Practice and Research. 2007.
2. Peterson-Falzone SJ, Hardin-Jones MA, Karnell MP: Cleft palate speech: Mosby St. Louis; 2001.
3. Gooch JL, Hardin-Jones M, Chapman KL, Trost-Cardamone JE, Sussman J: Reliability of listener transcriptions of compensatory articulations. Cleft Palate Craniofac J. 2001;38(1):59-67.
4. Howard SJ, Heselwood BC: Learning and teaching phonetic transcription for clinical purposes. Clin Linguist Phon. 2002;16(5):371-401.
5. Ramsdell HL, Oller, DK, C.A. Ethington, CA: Predicting phonetic transcription agreement: insights from research in infant vocalizations. Clin Linguist Phon. 2007; 21(10): 793-831.
6. Bellis TH, Wohlgemuth B: The Incidence of Cleft Lip and Palate Deformities in the South-east of Scotland (1971-1990). Br J Orthod. 1999;26(2):121-5. doi: 10.1093/ortho/26.2.121.
7. Harding A, Grunwell P: Active versus passive cleft-type speech characteristics. Int J Lang Commun Disord. 1998;33(3):329-52.
8. Trost JE: Articulatory additions to the classical description of the speech of persons with cleft palate. The Cleft Palate J. 1981;18(3):193-203.
9. Lohmander A, Willadsen E, Persson C, Henningsson G, Bowden M, Hutters B: Methodology for speech assessment in the Scandcleft project—An international randomized clinical trial on palatal surgery: Experiences from a pilot study. Cleft Palate Craniofac J. 2009;46(4):347-62. doi: 10.1597/08-039.1.
10. Gibbon FE, Crampin LB: An electropalatographic investigation of middorsum palatal stops in an adult with repaired cleft palate. Cleft Palate Craniofac J. 2001;38(2):96-105.
11. Munson B, Edwards J, Schellinger SK, Beckman ME, Meyer MK: Deconstructing phonetic transcription: covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. Clin Linguist Phon. 2010;24(4-5):245-60. doi: 10.3109/02699200903532524.
12. Sell D: Issues in perceptual speech analysis in cleft palate and related disorders: a review. Int J Lang Commun Disord. 2005;40(2):103-21. doi: 10.1080/13682820400016522.
13. Kuehn DP, Moller KT: Speech and language issues in the cleft palate population: the state of the art. Cleft Palate Craniofac J. 2000;37(4):1-35.
14. Howard S, Heselwood B: Instrumental and perceptual phonetic analyses: The case for two-tier transcriptions. Clin Linguist Phon. 2011;25(11-12):940-8.

15. Roxburgh Z, Cleland J, Scobbie JM: Multiple phonetically trained-listener comparisons of speech before and after articulatory intervention in two children with repaired submucous cleft palate. Clin Ling Phon. 2016;30(3-5):398-415. doi: 10.3109/02699206.2015.1135477.

16. Klinto K, Lohmander A: Does the recording medium influence phonetic transcription of cleft palate speech? Int J Lang Commun Disord. 2017;52(4):440-9. doi: 10.1111/1460-6984.12282.

17. Forrest K, Weismer G, Hodge M, Dinnsen DA, Elbert M: Statistical analysis of word-initial/k/and/t/produced by normal and phonologically disordered children. Clin Linguist Phon. 1990;4(4):327-40.

18. Maxwell EM, Weismer G: The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. Appl Psycholinguist. 1982;3(1):29-43.

19. Gibbon F: Bibliography of electropalatographic (EPG) studies in English (1957–2013). Dept Speech Hear Sci, Univ College Cork, Ireland, Rep Staeno. 2013:05-21.

20. Cleland J, Scobbie JM, Heyde C, Roxburgh Z, Wrench AA: Covert contrast and covert errors in persistent velar fronting. Clin Linguist Phon. 2017;31(1):35-55.

21. Gibbon FE: Undifferentiated lingual gestures in children with articulation/phonological disorders. J Speech Lang Hear Res . 1999;42(2):382-97.

22. Cleland J, Scobbie JM, Roxburgh Z, Heyde C, Wrench A: Enabling new articulatory gestures in children with persistent speech sound disorders using ultrasound visual biofeedback. J Speech Lang Hear Res. In Press.

23. Hardcastle WJ, Gibbon F: Electropalatography and its clinical applications. Instrumental Clinical Phonetics. 1997:149-93.

24. Gibbon FE: Abnormal patterns of tongue-palate contact in the speech of individuals with cleft palate. Clin Linguist Phon. 2004;18(4-5):285-311. doi: 10.1080/02699200410001663362.

25. Bernhardt MB, Bacsfalvi P, Adler-Bock M, Shimizu R, Cheney A, Giesbrecht N, O'connell M, Sirianni J, Radanov B: Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. Clin Linguist Phon. 2008; 22(2):149-162.

26. Heyde CJ, Scobbie JM, Lickley R, Drake EK: How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound. Clin Linguist Phon. 2016;30(3-5):292-312.

27. McAllister Byun T, Buchwald A, Mizoguchi A: Covert contrast in velar fronting: An acoustic and ultrasound study. Clin Linguist Phon. 2016;30(3-5):249-76.

28. Bressmann T, Radovanovic B, Kulkarni GV, Klaiman P, Fisher D: An ultrasonographic investigation of cleft-type compensatory articulations of voiceless velar stops. Clinical Linguist Phon. 2011;25(11-12):1028-33.

29. Lee A, Gibbon FE, Crampin L, Yuen I, McLennan G: The national CLEFTNET project for individuals with speech disorders associated with cleft palate. Adv Speech Lang Pathol. 2007;9(1):57-64.

30. Debbie S, Anne H, Pamela G: GOS.SP.ASS.'98: an assessment for speech disorders associated with cleft palate and/or velopharyngeal dysfunction (revised). Int J Lang Commun Disord. 1999;34(1):17-33. doi: 10.1080/136828299247595.

31. Articulate Instruments Ltd: Articulate Assistant Advanced Ultrasound Module user manual, revision 2.14. Articulate Instruments, Edinburgh. 2014.

32. Cleland J, Wrench A, Lloyd S, Sugden E: ULTRAX2020: Ultrasound Technology for Optimising the Treatment of Speech Disorders: Clinicians' Resource Manual. 2018.

33. Ladefoged P. The revised international phonetic alphabet. Language. 1990;66(3):550-2.

34. Ball MJ, Howard SJ, Miller K. Revisions to the extIPA chart. J Int Phon Assoc. 2018;48(2):155-64.

35. Shriberg LD, Lof GL: Reliability studies in broad and narrow phonetic transcription. Clin Linguist Phon. 1991;5(3):225-79. doi: 10.3109/02699209108986113.